IT Unit 3

Topic 2

# Ch. 2, Data Analytics: Drawing Conclusions, Part 1

**Integrity of data,** p 111- 121

1. Why is data integrity important? Explain possible effects of using data that lacks integrity.

* For a computer to produce useful information, the data that is input into a database must have integrity. Whether you are storing transmitting or archiving data, you must be sure that its integrity is maintained.
* A lack of integrity will affect your research to the point where your data isn’t accurate

**Timeliness**

1. What does timeliness refer to? Give some examples of data that is not timely.

* Data must be timely for it produce unable information, that’s means it needs to be processed while it is current and there should be no significant, delays in retrieving it. Using older data may not be accurate to current research requirements.

**Authenticity**

1. What are the characteristics of authentic data?

* Comes from the author and/or source it claims to be from
* Has not been deliberately corrupted
* Is not fake or disguised as something else
* Has not been changed without authorisation
* Is what it claims to be and does not misrepresent itself
* Does not aim to mislead or deceive not pretending to anything else
* Does not lie

1. List some challenges to the authenticity of data.

* Fake identify use on online
* Torrent websites have increase the likelihood of downloading data that differs from original
* People who upload cracked and corrupted files

1. Authenticity techniques. How can you authenticate data, both digital and non-digital?

* Digital use either a type of code or some type of validation usually email. Using codes than will only be receive by the product has been purchase either physically or digital
* Non-digital compare original documents with any allegedly accurate copies. Contact the original authors of documents to verify their authenticity. If that’s possible.

**Relevance**

1. What makes data relevant?

* If the content of the data shares the same characteristics as data already collected it should be relevant but the data doesn’t share the same features, it won’t be useful to the current purpose of the data.

**Accuracy**

1. Distinguish between content and form in terms of accuracy of data.

* There are two main characteristics of data accuracy: **Content** (**Functionality**)and **form** (**appearance**). Content is divided into two parts correctness and completeness

1. Briefly elaborate on the following challenges to data accuracy
   1. Correctness

* Correctness means that the values stored for a given object must be correct
  1. Completeness
* Completeness means that your data set is just that: complete.
  1. Clarity
* The form of data is important as well as the content because it will remove ambiguity about the content.
  1. Consistency
* Correct, unambiguous data can still cause a problem in a database it is not consistent, inconsistent data is unwelcome because it means the data is reliable.

1. Measures to improve accuracy, briefly elaborate on the following measures;
   1. Correctness

* You can perform data quality assurance to cleanse or scrub data. This will identify and remove or repair data that is incomplete, inaccurate, irrelevant or inconsistent.
  1. Completeness
* For data that you suspect has aps because it has been collected but not published in full, you could contact the original data collectors to ask if they have unpublished data that they could provide you
  1. Clarity
* Enforce data formatting and validation rules in your spreadsheets and database that prevents misinterpretation of the data.
  1. Consistency
* When formatting questions that deal directly with people, ask the same question again in different ways to check the answer consistency. This is important for questions about sensitive issues.